

WHAT IS CLAIMED IS:

1. A method of identifying a caller of a call from the caller to a recipient, the method comprising:

- (a) receiving a voice input from the caller;
- (b) applying characteristics of the voice input to a plurality of acoustic models, which comprises a generic acoustic model and acoustic models of any previously identified callers, to obtain a plurality of respective acoustic scores;
- (c) identifying the caller as one of the previously identified callers or a new caller based on the plurality of acoustic scores; and
- (d) if the caller is identified as a new caller in step (c), generating a new acoustic model for the new caller, which is specific to the new caller.

2. The method of claim 1 wherein identifying in (c) comprises:

- (c)(1) identifying the caller as one of the previously identified callers if the acoustic score for the respective acoustic model is better than the acoustic score for the generic acoustic model; and
- (c)(2) identifying the caller as a new caller if the acoustic score for the generic acoustic model is better than the acoustic scores

for the acoustic models of the plurality of previously identified callers.

3. The method of claim 1 wherein:

step (a) comprises segmenting the voice input into a sequence of recognized speech units using the generic acoustic model;

each of the plurality of acoustic models comprises models of the speech units segmented in step (a); and

step (b) comprises applying the characteristics of the voice input to a sequence of the models of the speech units segmented in step (d) for the plurality of acoustic models.

4. The method of claim 1 wherein each of the plurality of acoustic models comprises models of speech units and wherein the method further comprises:

(e) if the caller is identified as one of the previously identified callers in step (c), updating the respective acoustic model for the previously identified caller by modifying the models of the speech units that are included in the voice input, based on the characteristics of the voice input.

5. The method of claim 4 wherein step (e) comprises modifying the models of the speech units that are

included in the voice input based on as little as a single utterance.

6. The method of claim 1 and further comprising:

(e) storing the new acoustic model in an acoustic model repository with the plurality of acoustic models such that the new acoustic model becomes one of the plurality of acoustic models in step (b) and the new caller is included as a previously identified caller.

7. The method of claim 1 wherein the generic acoustic model comprises caller-independent models of a plurality of speech units, and wherein step (d) comprises:

(d)(1) generating the new acoustic model from the caller-independent models of the generic acoustic model and modifying the caller-independent models of the speech units that are included in the voice input to represent the characteristics of the voice input received from the new caller.

8. The method of claim 1 wherein steps (a) through (c) are performed without alerting the caller during the call that the caller is being identified.

9. The method of claim 1 wherein:

step (b) comprises splitting the voice input into subsections and applying the characteristics of each subsection to the plurality of acoustic models to obtain a plurality of respective acoustic scores that represent how well the characteristics in each subsection match the respective acoustic models; and

step (c) comprises, for each subsection, identifying the acoustic model having the best acoustic score for that subsection and identifying the caller as one of the previously identified callers only if the best acoustic scores for all subsections correspond to the same previously identified caller.

10. The method of claim 1 and further comprising:

- (e) maintaining a caller-specific language model for each of the previously identified callers based on the voice inputs from those callers;
- (f) applying the characteristics of the voice input to the generic acoustic model and each of the caller-specific language models to produce a plurality of recognized speech unit sequences;
- (g) choosing the recognized speech unit sequence that has a highest probability

relative to probabilities of the other recognized speech unit sequences; and

- (h) identifying the caller based at least in part on the recognized speech unit sequence having the highest probability.

11. The method of claim 10 and further comprising:

- (i) if the caller identified in step (h) is different than the caller identified in step (c), generating a user prompt for a manual review of at least one of the following: the voice input, the recognized speech unit sequence, the identified callers, the acoustic model of the caller identified in step (c), and of the caller-specific language model of the caller identified in step (h).

12. The method of claim 1 and further comprising:

- (e) using a distance measure between the plurality of acoustic models of the previously identified callers to flag certain acoustic models for merging together.

13. The method of claim 12 wherein step (e) comprises flagging the certain acoustic models for manual inspection.

14. A system for identifying a caller of a call from the caller to a recipient, the system comprising:

a receiver for receiving a voice input from the caller;

an acoustic model repository comprising a plurality of acoustic models, including a generic acoustic model and acoustic models of any previously identified callers;

means for applying characteristics of the voice input to the plurality of acoustic models to produce a plurality of respective acoustic scores;

means for identifying the caller as one of the previously identified callers or a new caller based on the plurality of acoustic scores; and

acoustic model generator means for generating a new acoustic model for the new caller if the acoustic score for the generic acoustic model is better than the acoustic scores for the acoustic models of the plurality of previously identified callers.

15. The system of claim 14 and further wherein:

the system further comprises a speech recognizer, which segments the voice input into a sequence of recognized speech units using the generic acoustic model;

each of the plurality of acoustic models comprises models of the speech units recognized by the speech recognizer; and the means for applying comprises means for applying the characteristics of the voice input to a sequence of the models of the speech units segmented by the speech recognizer for the plurality of acoustic models.

16. The system of claim 14 wherein:

each of the plurality of acoustic models comprises models of speech units; and the system further comprises an acoustic model updating module, which if the caller is identified as one of the previously identified callers, updates the respective acoustic model for the previously identified caller by modifying the models of the speech units that are included in the voice input, based on the characteristics of the voice input.

17. The system of claim 16 wherein the acoustic model updating module is capable of modifying the models of the speech units that are included in the voice input based on as little as a single utterance from the caller.

18. The system of claim 14 wherein the acoustic model generator means stores the new acoustic model in the acoustic model repository such that the new acoustic model becomes one of the plurality of acoustic models and the new caller is included as a previously identified caller.

19. The system of claim 18 wherein:

the generic acoustic model comprises caller-independent models of a plurality of speech units; and

the acoustic model generator generates the new acoustic model from the caller-independent models of the generic acoustic model and modifies the caller-independent models of the speech units that are included in the voice input to represent the characteristics of the voice input received from the new caller.

20. The system of claim 14 wherein system is configured to receive the voice input and identify the caller without alerting the caller during the call that the caller is being identified.

21. The system of claim 14 wherein:

the means for applying comprises means for splitting the voice input into subsections and applying the characteristics of each subsection to the plurality of acoustic



models to obtain a plurality of respective acoustic scores that represent how well the characteristics in each subsection match the respective acoustic models; and  
the means for identifying comprises, for each subsection, means for identifying the acoustic model having the best acoustic score for that subsection and means for identifying the caller as one of the previously identified callers only if the best acoustic scores for all subsections correspond to the same previously identified caller.

22. The system of claim 14 and further comprising:

a language model repository for storing a caller-specific language model for each of the previously identified callers based on the voice inputs from those callers;

means for applying the characteristics of the voice input to the generic acoustic model and each of the caller-specific language models to produce a plurality of recognized speech unit sequences; and

means for choosing the recognized speech unit sequence that has a highest probability relative to probabilities of the other recognized speech unit sequences, wherein the means for identifying identifies the caller based at least in part on the

recognized speech unit sequence having the highest probability.

23. The system of claim 22 wherein the means for identifying comprises means for generating a user prompt for a manual review of at least one of the following: (1) the voice input, the recognized speech unit sequence having the highest probability, (2) the caller-specific language model producing the recognized speech unit sequence having the highest probability, and (3) the acoustic model having the best acoustic score, if the caller-specific language model having the highest probability corresponds to a different caller than the acoustic model having the best acoustic score in (3).

24. The system of claim 14 and further comprising:  
means for flagging certain acoustic models for merging together based on a distance measure between the plurality of acoustic models.

25. The method of claim 24 wherein the means for flagging comprises means for flagging the certain acoustic models for manual inspection.

26. A computer-readable medium comprising computer-executable instructions that, when executed by a computer, performs the method comprising:

- (a) receiving a voice input of a call from a caller;
- (b) applying characteristics of the voice input to a plurality of acoustic models, which comprises a generic acoustic model and acoustic models of any previously identified callers, to obtain a plurality of respective acoustic scores that represent how well the characteristics match the respective acoustic models; and
- (c) identifying the caller as one of the previously identified callers or a new caller based on the plurality of acoustic scores; and
- (d) if the caller is identified as a new caller in step (c), generating a new acoustic model for the new caller, which is specific to the new caller.

27. The computer-readable medium of claim 26 wherein:

- step (a) comprises segmenting the voice input into a sequence of recognized speech units using the generic acoustic model;
- each of the plurality of acoustic models comprises models of the speech units segmented in step (a); and
- step (b) comprises applying the characteristics of the voice input to a sequence of the models of the speech units segmented in

step (a) for the plurality of acoustic models.

28. The computer-readable medium of claim 26 wherein each of the plurality of acoustic models comprises models of speech units and wherein the method further comprises:

- (e) if the caller is identified as one of the previously identified callers in step (c), updating the respective acoustic model for the previously identified caller by modifying the models of the speech units that are included in the voice input, based on the characteristics of the voice input.

29. The computer-readable medium of claim 26 wherein the method further comprises:

- (e) storing the new acoustic model in an acoustic model repository with the plurality of acoustic models such that the new acoustic model becomes one of the plurality of acoustic models in step (b) and the new caller is included as a previously identified caller.

30. The computer-readable medium of claim 29 wherein the generic acoustic model comprises caller-independent models of a plurality of speech units, and wherein step (d) comprises:

- (d)(1) generating the new acoustic model from the caller-independent models of the generic acoustic model and modifying the caller-independent models of the speech units that are included in the voice input to represent the characteristics of the voice input received from the new caller.

31. The computer-readable medium of claim 26 wherein the method further comprises:

- (e) maintaining a caller-specific language model for each of the previously identified callers; and
- (f) identifying the caller based at least in part on probabilities of recognized speech unit sequences produced by the caller-specific language models from the voice input.

32. The computer-readable medium of claim 31 wherein the method further comprises:

- (g) if the caller identified in step (f) is different than the caller identified in step (c), generating a user prompt for a manual review of at least one of the following: the voice input, the recognized speech unit sequence, the identified callers, the acoustic model of the caller identified in step (c), and of the caller-

specific language model of the caller identified in step (f).

33. The computer-readable medium of claim 26 wherein the method further comprises:

- (e) using a distance measure between the plurality of acoustic models of the previously identified callers to flag certain acoustic models for merging together.

34. The computer-readable medium of claim 33 wherein step (e) comprises flagging the certain acoustic models for manual inspection.

35. A method of identifying a caller of a call from the caller to a recipient, the method comprising:

- (a) receiving a voice input from;
- (b) segmenting the voice input into a sequence of recognized speech units using a caller-independent, generic acoustic model;
- (c) applying characteristics of the voice input to a sequence of speech unit models of the recognized speech units within a plurality of acoustic models, which comprises the generic acoustic model and acoustic models of any previously identified callers; and
- (d) identifying the caller as one of a plurality of previously identified callers or as a new caller based on how well the

characteristics of the voice input fit the plurality of acoustic models.

36. The method of claim 35 and further comprising:
- (e) if the caller is identified as a new caller in step (d), generating a new acoustic model for the new caller from the generic acoustic model by modifying the speech unit models of the speech units that are included in the voice input to represent the characteristics of the voice input received from the new caller; and
  - (f) storing the new acoustic model in an acoustic model repository with the plurality of acoustic models such that the new acoustic model becomes one of the plurality of acoustic models in step (c) and the new caller is included as a previously identified caller.
37. The method of claim 35 and further comprising:
- (e) maintaining a caller-specific language model for each of the previously identified callers based on the voice inputs from those callers;
  - (f) applying the characteristics of the voice input to the generic acoustic model and each of the caller-specific language models to produce a plurality of recognized speech unit sequences;

- (g) choosing the recognized speech unit sequence that has a highest probability relative to probabilities of the other recognized speech unit sequences;
- (h) identifying the caller based on the recognized speech unit sequence having the highest probability; and
- (i) if the caller identified in step (h) is different than the caller identified in step (d), generating a user prompt for a manual review of at least one of the following: the voice input, the recognized speech unit sequence, the identified callers, the acoustic model of the caller identified in step (d), and of the caller-specific language model of the caller identified in step (h).

38. The computer-readable medium of claim 35 wherein the method further comprises:

- (e) using a distance measure between the plurality of acoustic models of the previously identified callers to flag certain acoustic models for merging together.